Journal of Applied Psychology
1986, Vol. 71, No. 2, 232-245

0021-9010/86/$00.75

A Cumulative Study of the Effectiveness of Managerial Training

Michael J. Burke
Department of Management
New York University

Russell R. Day
Illinois Institute of Technology and
Chicago and North Western Transportation Company

The published and unpublished literature on the effectiveness of managerial training has produced conflicting results and left more unanswered questions than definitive statements concerning the effectiveness of managerial training. In the present study, meta-analysis procedures were applied to the results of 70 managerial training studies to empirically integrate the findings of the studies. The meta-analysis results for 34 distributions of managerial training effects representing six training content areas, seven training methods, and four types of criteria (subjective learning, objective learning, subjective behavior, and objective results) indicated that managerial training is, on the average, moderately effective. For 12 of the 17 managerial training method distributions, the 90% lower bound credibility values were positive, and thus the effectiveness of these training methods, at least at a minimal level, can be generalized to new situations. It is stressed that although this meta-analysis assisted in clarifying what we have learned about managerial training, a great deal of empirical research on managerial training is needed before more conclusive statements can be made.

Every year, managerial training and development programs are implemented in most private and public organizations. The pervasiveness of managerial training is well recognized (cf. Lundberg, Dunbar, & Bayless, 1973; Wexley & Latham, 1981; Wikstrom, 1973). The objective of most managerial training and development programs is to teach or improve various managerial skills to improve on-the-job performance (Goldstein, 1980; Wexley & Latham, 1981). In particular, a great deal of managerial training is focused on improving job performance in the areas of human relations, self-awareness, problem solving and decision making, motivation/values, and general management. Many organizations, however, are unaware of the effectiveness of training programs in these areas. This lack of knowledge concerning the results of managerial training is primarily due to the lack of evaluative research on these programs. Four notable reviews (Campbell, 1971; Campbell, Dunnette, Lawler, & Weick, 1970; Goldstein, 1980; Wexley, 1984) attempted to integrate the available evaluative managerial training research. Campbell et al. (1970) concluded that the literature on the relation of training to management performance reveals little about what kind of knowledge and skills contribute to managerial effectiveness. According to Campbell (1971), "the training development literature is voluminous, nonempirical, nontheoretical, poorly written, and dull" (p. 565). Unfortunately, as in his review and the Campbell et al. (1970) review, Goldstein's (1980) review indicated that the vast majority of literature in the managerial training area, and training in general, was still not empirical or based on theory. A large proportion of the managerial training and development literature was dominated by low-utility anecdotal presentations (Goldstein, 1980). Wexley (1984) concluded that considerably more research is urgently needed in the area of management development. In general, these four reviews illustrated that the relation between managerial training and the acquisition of managerial skills is not clear.

Although noteworthy, these reviews left some important questions unanswered. First, How effective, in general, is managerial training with respect to different criteria? In addition, there are many different types (i.e., content areas) of managerial training. As noted above, these include human relations, general management functions, problem solving and decision making, self-awareness, motivation/values, and specialties. Which types of management training programs are effective and to what degree? Moreover, there are numerous managerial training methods (e.g., lecture, discussion, role playing). Important questions remain concerning the relative effectiveness of different methods or combinations of methods in improving learning or acquisition of skills. In order to obtain answers to these questions, the results of management training studies must be empirically integrated.

Integrating the results of managerial training studies, as well as studies for other organizational interventions, via meta-analytic techniques (cf. Glass, McGaw, & Smith, 1981; Hunter, Schmidt & Jackson, 1982) will assist in determining the effectiveness of interventions. The findings of such research are likely to be of theoretical import to researchers as well as practical import to organizational decision makers. The purpose of this study was to apply the process of meta-analysis (cf. Glass et al., 1981; Hunter et al., 1982) to the available managerial training and development literature to determine the effectiveness of managerial training.

We attempted to answer the following questions: (a) Across studies with respect to various criteria, how effective is managerial training? (b) For each type of criterion measure, what is the relative effectiveness of different types of managerial training (i.e., human relations, general management, self-awareness, problem solving and decision making, motivation/values, and specialties)?

The authors would like to thank Nambury Raju, Frank Schmidt, and two anonymous reviewers for helpful comments on an earlier draft of this article.

Correspondence concerning this article should be addressed to Michael J. Burke, 600 Tisch Hall, Department of Management, New York University, 40 West 4th Street, New York, New York 10003.

(c) For each type of criterion measure, what is the relative effectiveness of different managerial training methods and combinations of methods?

Method

Selection of Studies and Coding of Study Characteristics

Each study in the present analysis had to meet three criteria: (a) The study involved managerial or supervisory personnel, (b) it evaluated the effectiveness of one or more training programs, and (c) it included at least one control or comparison group. Studies were located by conducting computer searches of the ERIC and PsychINFO indexes and by scanning bibliographies of the published and unpublished sources. The degree of methodological rigor or quality of the research design was not a selection criterion. After all literature searches, the total number of nonredundant papers collected for analysis was 70. Subsequent to the identification and selection of studies was the task of coding each study for analysis.

Study characteristics such as sample size and criterion reliability were collected and coded in order to correct reported results for artifacts such as sampling error and attenuation. An attempt was also made to include all pertinent information related to managerial training research. The studies were coded for the content area of training (human relations, problem solving and decision making, and so forth), the type of training method (technique) used, the type of outcome (criterion) variable measured, the type of industry in which the study was performed, length of training time, time between training and evaluation, assignment of subjects, type of comparisons made, use of before and after measures, and whether the results reported were significant or not. Subject characteristics included in the coding were sex, level of management, and years of experience as a manager or supervisor. In addition, type of trainer (i.e., professional trainer, researcher, other) was included in the coding. Descriptions of all the studies are presented by training content area in Table 1.

In order to ensure the reliability of the coding scheme developed for this meta-analysis, two studies were randomly selected and independently read and coded by five judges. The judges were two doctoral-level industrial psychology students and three industrial/organizational psychologists. The pairwise interrater agreement ranged from 87% to 90%. The overall interrater agreement was 88%.

Definitions of Training Content Areas, Training Methods, and Criteria

In categorizing studies, the following definitions (descriptions) were used for training content. They are similar to those developed by Campbell et al. (1970).

General management programs. This is the broadest type of development effort and typically includes material on labor relations, management theory and practice, company policies and procedures, labor economics, and general management functions. The primary goal of these training programs is the teaching of facts, concepts, and skills.

Human relations/leadership programs. The content of these programs is narrower than that of the general management programs category in that the focus is on human relations problems of leadership, supervision, attitudes toward employees, and communication.

Self-awareness programs. The content of these programs is on understanding one's own behavior and how one's behavior is viewed by others, identifying the so-called games people play, and learning about one's strengths and weaknesses. Typical training methods are sensitivity training, laboratory training, T-groups, and transactional analysis.

Problem-solving/decision-making programs. The emphasis of these programs is on teaching generalized problem-solving and decision-making skills that would be applicable to a wide range of work problems that managers encounter.

Rater training programs. In these programs managers are trained to minimize errors when they are observing and evaluating their subordinates.

Motivation/values training programs. The content of these programs deals with increasing managers' motivation or modifying a manager's values or attitudes.

A set of definitions was also used for categorizing studies by training method. These training method definitions were as follows.

Lecture. This is the most traditional method for presenting information, usually involving a carefully prepared oral presentation on a subject by a qualified individual (Reith, 1976).

Lecture/group discussion. This category includes the lecture method described above, as well as group discussion. The discussion often focuses on specific problems or new areas of knowledge. Active participation by the trainee is encouraged. This type of training provides the opportunity for clarification, feedback, and considerable trainee involvement.

Leader Match. This is a method used in human relations/leadership training that incorporates self-paced workbooks and standardized problems (Fiedler, Chemers, & Mahar, 1976). The self-paced workbooks are designed to teach trainees how to apply the principles of Contingency Theory (Fiedler, 1964, 1967) to their leadership position. More specifically, the Leader Match training program is designed to teach individuals how to become aware of their leadership style through the self-paced workbook, how to diagnose the situation in which they work, and how to change the situation to fit their personality.

Sensitivity training. The classical model of sensitivity training is a group meeting without an agenda in which participants discuss topics dealing with the "here and now" of the group process (Hinrichs, 1976). Sensitivity training represents both a distinct training method and a specific content area (defined above as self-awareness). Laboratory education, a label applied to a more complete training program in which the main component is some form of sensitivity group, was considered sensitivity training for this study.

Behavioral modeling. This method of training is based on Bandura's (1977) social learning theory and emphasizes observation, modeling, and feedback in order to modify behavior.

Lecture/group discussion with role playing or practice. This category represents a common combination of methods using lecture/group discussion as well as role playing or practice. Role playing is designed such that trainees assume parts of specific personalities in a contrived situation. It typically involves modeling, practice, and feedback. Practice refers to providing the trainee with the opportunity to engage in the activity or skill which is being taught.

Multiple techniques. This category includes studies that employed three (excluding the Lecture/group discussion with role playing category) or more training methods. Most of these studies had a common method such as lecture or group discussion and more specific methods such as a case study, business game, or audiovisual presentation.

We developed four criterion-measure categories on the basis of two dimensions: (a) level of criterion and (b) subjectivity–objectivity. The level of criterion dimensions was determined according to Kirkpatrick (1976); three of Kirkpatrick's four levels were used: learning, behavior, and results.

Subjective learning. This category included measures that assessed what principles, facts, attitudes, and skills were learned during or by the end of training as communicated in statements of opinion, belief, or judgment completed by the trainee or trainer (observer).

Objective learning. Included in this category were measures that assessed what principles, facts, attitudes, and skills were learned during or by the end of training by objective means, such as number of errors made or number of solutions reached, or by standardized test (e.g., knowledge test).

Subjective behavior. This category included measures that evaluated changes in on-the-job behavior perceived by trainees, peers, or supervisor.

(*text continues on page 237*)

Table 1
Managerial Training Studies Used in Meta-Analysis

Author(s)	Sample description	Sample size	Type of training	Criteria
Human relations/leadership				
Baum, Sorensen, & Place (1970)	Insurance company middle managers	12-T[a] 12-C	Lecture, small group discussion, case studies	Perceptions of influence exercised by subordinates
Burnaska (1976)	General Electric middle managers	62-T 62-C	Behavioral modeling	Ratings of interpersonal skills
Byham, Adams, & Kiggins (1976)	Financial accounting supervisors	8-T 8-C	Behavioral modeling, practice, audiovisual aids	Ratings on interpersonal skills from subordinates
Canter (1951)	Insurance company supervisors	18-T 18-C	Lecture	Knowledge tests, questionnaires on human relations
Carron (1964)	Chemical company supervisors	23-T 12-C	Group discussion, role playing	Leadership Opinion Questionnaire, Authoritarianism Scale
Csoka & Bons (1978)	College ROTC military leaders	26-T1[b] 27-C1 38-T2 76-C2	Self-paced workbook, Leader Match	T1. Composite leadership ratings from superiors and peers T2. Rank orderings by superiors
Decker (1982)	First-line hospital supervisors	12-T 12-C	Behavioral modeling, audiovisual aids	Ratings by trained raters of ability to coach and handle employee complaints
DiVesta (1954)	Air Force medical administrative supervisors	94-T	Lecture	Knowledge tests, performance ratings
Fiedler & Mahar (1979a)				
Study 1	Public health volunteer leaders	9-T 27-C 11-T 33-C	Self-paced workbook, Leader Match	Performance ratings by superiors
Study 2	Same	18-T 12-C 16-T 10-C	Leader Match vs. alternative training program of similar format and length vs. control group	Same
Fiedler & Mahar (1979b)	College ROTC military leaders	190-T 215-C	Leader Match	Overall performance ratings by advisors and peers
Fiedler, Mahar, & Schmidt (1976)				
Study A	Urban county government middle managers	15-T 15-C	Leader Match, group discussion, audiovisual aids	Biyearly performance evaluations (ratings)
Study B	Police	7-T 10-C	Same	Performance ratings by immediate superior
Study C	Public works supervisors	14-T 13-C	Same	Performance ratings by two superiors
Fleishman (1953)	Industrial foremen	90-T 32-C	Lecture, group discussion, role playing	Supervisory Behavior Description Questionnaire, Foreman's Leadership Opinion Questionnaire
Goldstein & Sorcher (1974)	General Electric foremen	100-T 100-C	Behavioral modeling	Subordinate productivity
Hand & Slocum (1970)/ Hand, Richard, & Slocum (1973)[c]	Steel plant middle managers	21-T 21-C	Group discussion, case studies, practice	Leadership Opinion Questionnaire, Supervisory Behavior Description Questionnaire, knowledge test, sensitivity to others questionnaire, performance ratings, salary increases
Harris & Fleishman (1955)	First-level supervisors, truck manufacturing	39-T 59-C	Lecture, group discussion, visual aids	SBD ratings
Jennings (1954)	First-level supervisors	20-T 20-C	Case study, forced leadership method	Performance rankings
Latham & Saari (1979)	Paper company foremen	20-T 20-C	Behavioral modeling	Knowledge tests, ratings by superiors, ratings during role play
Lawshe, Bolda, & Brune (1959)				
Study 5	First-level supervisors	12-T 15-C	Role playing	Ratings in performance in skills test

Table 1 (*continued*)

Author(s)	Sample description	Sample size	Type of training	Criteria
Leister, Borden, & Fiedler (1977)	Naval officers	27-T 29-C	Leader Match, visual aids	Performance ratings by superiors
Maier (1953)	First-level supervisors	44 T groups 36 C groups	Role playing, group discussion	Ratings during role play, number of solutions reached
Mayo & DuBois (1963)	Military officers	211-T 211-C	Lecture	Ratings by superiors
Moon & Hariton (1958)	General Electric engineering managers	66-T 67-C	Lecture, group discussion, role playing, practice	Survey ratings by subordinates
Mosel & Tsacnaris (1954)	Air Force officers	83-T 44-C	Lecture	How Supervise
Moses & Ritchie (1976)	Telephone company supervisors	90-T 93-C	Behavioral modeling	Ratings during role play tests on interpersonal skill
Smith (1976)				
Study 1	IBM middle managers	18-T 13-C	Behavioral modeling	Subordinate ratings
Study 2	Same	36-T	Behavior modeling and team building vs. behavioral modeling vs. lecture–discussion	Skills test in communication, sales performance
Stroud (1959)	Telephone company supervisors	103-T 91-C	Lecture, group discussion, role playing	Superiors ratings of performance, Leader, Behavior Description Questionnaire
Trice & Belasco (1968)	First-level supervisors	55-T 55-C	Lecture, group discussion, role playing	Knowledge test, attitude scale
General management functions				
Alves & Hardy (1963)	Los Angeles County supervisors	46-T 46-C	Lecture, case studies	Attitude questionnaire from subordinates
Buchanan & Brunstetter (1959)	First-level supervisors	224-T	Laboratory education	Survey questionnaire
Couch & Strother (1971)	First-level manufacturing supervisors	30-T 30-C	Lecture	Self-reported critical incidents
Goodacre (1957)	B. F. Goodrich middle managers	400-T 400-C	Lecture, group discussion	Knowledge tests, attitude survey, performance ratings
Goodacre (1963)	General Electric middle managers	218-T 279-C	Lecture, practice	Questionnaires reported by subordinates
House & Tosi (1963)	Engineer managers	24-T 33-C	Lecture	Self-report questionnaire
Ivancevich & Smith (1981)	Sales managers	20-T 20-T 20-C	Videotaping, role play vs. role play, lecture vs. controls	Goal setting rating forms reported by subordinates
Kondrasuk (1972)	Public service middle managers	13-T 13-C	Lecture, case study, role play, group discussion, audiovisual aids	Knowledge test
Latham & Kinne (1974)	First-level logging supervisors	10-T 10-C	Lecture, practice	Worker productivity
Mahar (1981)	School curriculum supervisors	5-T 5-C	Behavioral modeling, practice	Ratings of developed educational programs
Mahoney, Jerdee, & Korman (1960)	Middle managers	46-T 13-C	Conference and case study methods	Management Practices Quiz, case analysis, attitude scale
McGehee & Gardner (1955)	Textile production foremen	10-T 11-C	Lecture, practice	Knowledge test, attitude scale, performance ratings
Miner (1965)	Research & development supervisors	52-T 49-C	Lecture	Advancement, motivation needs
Schwarz, Stilwell, & Scanlan (1968)	Insurance company middle managers	28-T 29-C	Lecture, group discussion, case study, role play, audiovisual aids	Leadership Behavior Description Questionnaire by subordinates, self-reported critical incidents
Wexley & Nemeroff (1975)	Medical center supervisors	9-T 9-T	Role playing, practice, goal setting, feedback	Leadership Behavior Description Questionnaire by subordinates, subordinates absenteeism
Wilburn (1979)	First-level manufacturing	75-T 75-C	Lecture, conference, & case study method; role playing; practice	How Supervise

(*table continued*)

Table 1 (*continued*)

Author(s)	Sample description	Sample size	Type of training	Criteria
Wolfe & Moe (1973)	Health care supervisors	17-T 17-C	Lecture, case study	How Supervise
Problem solving/decision making				
Moffie, Calhoon, & O'Brien (1964)	Papermill middle managers	32-T 28-C	Standardized work problems	Watson-Glaser Critical Thinking Appraisal
Roy & Dolke (1971)	First-level supervisors	12-T 12-C	Lecture, case study, practice, group discussion	Knowledge test
Smith & Kight (1959)	First-level supervisors	33-T 69-C	Lecture, group discussion, role play, practice	Number of problems solved during test
Wilson, Mullen, & Morton (1968)	Middle managers	21-T 28-C	Conference & case study method, practice on standardized problems	Self-reported questionnaire
Self-awareness				
Argyris (1965)	Business executives	15-T 15-C	T-groups	Observer ratings during training
Boyd & Ellis (1963)	Public utility middle managers	42-T 12-T 10-C	T-groups vs. lecture-group discussion vs. no-training group	On-the-job behavior changes observed by subordinates
Bunker (1965)	Middle managers from various industries	229-T 112-C	T-groups	Open-ended behavior change descriptions by co-workers and self
Hand & Slocum (1972)	Middle managers	21-T 21-C	Lecture, laboratory education	Leadership Opinion Questionnaire, Supervisory Behavior Description Questionnaire, performance ratings, attitude scale
Harrison (1962)	Business executives	8-T 12-C	T-groups	Self-reported description of others
Harrison & Lubin (1965)	Middle managers	23-T 46-C	Laboratory education	Peer ratings for self-expression and amount of learning
Katz & Schwebel (1976)	Data processing	12-T 13-C	Laboratory education	Problem Analysis Questionnaire, peer ratings, Crowne-Marlowe Social Desirability Scale
Miles (1960 & 1965)	Elementary school principals	34-T 34-C	Laboratory education	Leader Behavior Description Questionnaire, ratings for behavior changes
Rubin (1967)	Supervisors from various occupations	30-T 11-C	T-groups	Sentence completion test for self acceptance and reduction in prejudice
Schutz & Allen (1966)	Middle managers	71-T 30-C	T-groups	Self-report questionnaire on interpersonal behavior
Smith (1963)	Middle managers	108-T 44-C	T-groups	Attitude scale
Steele (1968)	Middle managers	72-T 39-C	Laboratory education	Managerial Behavior Questionnaire on interpersonal values
Underwood (1965)	First-level supervisors	15-T 15-C	Laboratory education	Ratings by co-workers
Valiquet (1968)	Middle managers	34-T 15-C	Laboratory education	Description on open-ended questionnaire by peers
Motivation/values				
Arnoff & Litwin (1971)	Manufacturing executives	11-T 11-C	Lecture, group discussion, business games	Rate of advancement in company
Durand (1975)	Owners and operators of businesses	11-T 9-C	Lecture, case study, group discussion	Achievement motivation scores on Thematic Apperception Test, locus of control, interview, business activity questionnaire
Gruenfeld (1966)	Business executives	69-T 46-C	Lecture	Allport Vernon Study of Values
Miner (1960)	Research & development supervisors	55-T 30-C	Lecture	Sentence Completion Questionnaire
Viteles (1959)	Business executives	17-T 16-C	Lecture, group discussion	Knowledge test, Allport Vernon Study of Values
Rater training				
Ivancevich (1979)	Engineering supervisors	22-T 22-T	Workshop training vs. discussion & lecture	Reduction in halo and leniency rating errors

Table 1 (*continued*)

Author(s)	Sample description	Sample size	Type of training	Criteria
Levine & Butler (1952)	Manufacturing supervisors	11-T 9-T 9-C	Lecture vs. conference method vs. no training	Ability to make less biased performance ratings
Warmke & Billings (1979)	Nursing supervisors	14-T 10-T 15-T 13-C	Lecture vs. discussion vs. scale construction vs. no training	Reduction in performance rating errors

[a] T = trained group, C = comparison group.
[b] Indicates specific training group.
[c] Initial evaluations were reported in the 1970 article; follow-up analyses of the same study were reported in the 1973 article.

Objective results. This category included measures that evaluated tangible results, such as reduced costs, improved quality or quantity, promotions, and reduced number of errors in making performance ratings.

Meta-Analysis Procedure

Effect sizes were used as the common metric in this meta-analysis. Effect size or magnitude of effect can be defined as the normalized difference between a trained group and a comparison group. This can be restated as

$$ES = (M_t - \bar{M}_c)/S_w, \tag{1}$$

where ES is effect size, M_t is the treatment or trained group mean score on the dependent variable, M_c is the comparison or control group mean score on the dependent variable, and S_w is the within-group standard deviation on the dependent variable, given by the following formula:

$$S_w = \sqrt{\frac{(N_t - 1)S_t^2 + (N_c - 1)S_c^2}{N_t + N_c - 2}}, \tag{2}$$

where N_t is the number of subjects in the trained group, N_c is the number of subjects in the comparison group, S_t^2 is the variance for the trained group on the dependent variable, and S_c^2 is the variance for the comparison group on the dependent variable. Thus, an ES of +1 indicates that if a person at the mean of the comparison group were given the treatment (in this case training), he or she would be expected to rise to the 84th percentile of the comparison group.

An effect size was calculated on the outcome variable or variables reported in each study. If a study had more than one outcome variable, each variable was treated as independent and was entered separately into the overall analysis. For studies that reported only t, F, or other inferential statistics, ES was calculated by the appropriate conversion formulas (cf. Glass et al., 1981; Hunter et al., 1982; Schmidt, Hunter, & Pearlman, 1982). In addition, each ES was corrected for attenuation due to the unreliability of the criterion measure used in a particular study by the following equation:

$$ES_{ci} = \frac{ES_i}{\sqrt{r_{yyi}}}, \tag{3}$$

where ES_{ci} is the effect size corrected for criterion unreliability, ES_i is the attenuated effect size, and r_{yyi} is the criterion reliability. When a criterion reliability was not reported, the average value for the set of studies in an analysis was substituted.

Given that studies with larger sample sizes would provide more accurate estimates of the true or population effect size, each individual effect size was weighted on the basis of the specific sample size from which it came. The true mean effect size was thus calculated as

$$\bar{M}_\delta = \frac{\sum (N_i ES_{ci})}{\sum N_i}, \tag{4}$$

where $\bar{M}_\delta$ is the estimated true mean effect size, ES_{ci} is the unattenuated effect size for each outcome reported, and N_i is the total sample size used in calculating a specific ES_{ci} (cf. Hunter et al., 1982). In addition, the disattenuated effect size variance was calculated as

$$\hat{V}_{\delta c} = \frac{\sum N_i (ES_{ci} - \bar{M}_\delta)^2}{\sum N_i} \tag{5}$$

where $V_{\delta c}$ is the unattenuated effect size variance, ES_{ci} is the disattenuated effect size for each outcome reported, $\bar{M}_\delta$ is the mean population effect size, and N_i is the total sample size used in calculating a specific ES_{ci}.

In order to determine the true or population effect size variance, the sampling error across studies was calculated. The sampling error variance corrected for criterion unreliability was assessed by Hunter et al.'s (1982) formula:

$$V_e = \left(\frac{4(1 + \bar{M}_\delta^2/8)K}{\sum N_i}\right) \cdot \left(\frac{1}{K} \cdot \frac{1}{r_{yyi}}\right), \tag{6}$$

where V_e is the sampling error variance, r_{yyi} is the particular criterion reliability for the outcome reported, N_i is the total sample size, and K is the number of effect sizes reported.

The formula for estimating the population (true) effect size variance then became:

$$\hat{V}_\delta = \hat{V}_{\delta c} - V_e, \tag{7}$$

where $\hat{V}_\delta$ is the estimated true effect size variance, $\hat{V}_{\delta c}$ is the estimated disattenuated effect size variance, and V_e is the estimated sampling error variance.

The corrected variance term would be approximately zero if the effect size were constant across studies (i.e., homogeneous). If 75% or more of the observed effect size variance was accounted for by artifactual effects, then the hypothesis that the results were situation specific was rejected. Furthermore, when the difference was judged to be of practical significance (e.g., variance due to sampling error and criterion unreliability account for less than 75% of the observed variance), follow-up regression analyses were performed to determine what study characteristics other than sample size and criterion unreliability were contributing to the between-studies variance. In addition, after the unattenuated effect size variance was corrected for sampling error, lower bound credibility values were established for each estimated population effect size.

This procedure was followed in analyzing (a) training results for each of the four criterion-measure categories, (b) training results for each of the six training content areas with respect to the four criterion categories, and (c) training results for each of the seven training method categories, again with respect to the four criterion categories.

Results

Training Content

Subjective learning criteria. As shown in Table 2, only 12.4% of the observed variance over all studies using subjective learning

Table 2
Results for Content of Managerial Training on Subjective Learning Criteria

Training areas	No. of *ES*s	No. of studies	N	$\overline{ES}$	V_{obs}	V_e	% of variance accounted for	$\hat{M}_\delta$	$\hat{V}_\delta$	90% C.V.
Overall areas[a]	58	21	9,971	.31	.374	.028	12.4	.34	.199	−.23
Human relations	21	8	2,156	.69	.265	.051	31.9	.76	.108	.34
General management	20	4	6,878	.13	.378	.014	16.0	.14	.073	−.21
Self-awareness	15	7	832	.80	.524	.091	18.6	.86	.399	.06

Note. V_{obs} is the observed effect size variance. V_e is sampling error variance. $\overline{ES}$ is the observed weighted mean according to sample size.
[a] The content areas of problem solving, motivation/values and rater training were not included in separate breakdowns because only two or less studies in each area were found that used subjective learning criteria. However, they were included in the overall results for subjective learning criteria if data were available.

criteria was accounted for on the basis of artifactual effects. Human relations had the highest percentage of observed effect size variance accounted for (31.9%).

On the basis of 21 studies, the estimated true mean effect size for subjective learning criteria was .34. As Table 2 indicates, all of the true mean effect sizes for the training content areas were positive. Notably, the estimated true mean effect sizes for human relations and self-awareness were each more than three quarters of one standard deviation above the mean of the comparison groups. Moreover, the overall estimated true effect size variance for subjective learning criteria was .199 (see Table 2). For each of the training content areas, the estimated true effect size variances for subjective learning criteria were also substantially greater than zero.

Objective learning criteria. As indicated in Table 3, artifactual effects accounted for only 13.3% of the observed effect size variance over all studies for objective learning criteria. Likewise, artifactual sources of variance did not account for a substantial proportion of the observed effect size variance for three of the four content distributions (see Table 3). The results for the motivation/values content category should be treated with caution because only three studies, with a total of 12 effects, were found for this area.

The estimated true mean effect size and variance for objective learning criteria, over all studies, were .38 and .339, respectively. Consequently, a lower bound credibility value of −.37 was obtained. Similarly, negative lower bound credibility estimates were obtained for all other training content areas with the exception of motivation/values.

Subjective behavior criteria. The results for managerial training content distributions with studies that used subjective behavior criteria are presented in Table 4. Over all studies, using subjective measures of behavior as criteria, artifactual sources of variance only accounted for 13.8% of the observed effect size variance. Corresponding small percentages of variance accounted for by artifacts were found for general management and self-awareness training. Human relations training, however, had a relatively large percentage (60.5%) of the observed effect size variance accounted for by artifactual sources.

The estimated true mean effect size and variance for subjective behavior criteria, over all studies, were .49 and .344, respectively. As a result of these findings, a negative lower bound credibility value of −.26 was obtained for this analysis. General management and self-awareness training also had relatively large estimated true effect size variances and negative lower bound credibility values. On the other hand, human relations training had some variability in estimated true effect sizes (.061), yet the lower bound credibility value was positive (.12).

Objective results criteria. Over all studies using objective results as criteria, 22% of observed effect size variance was accounted for by artifacts (see Table 5). For the three content distributions, the corresponding percentages were higher: 31%, 33%, and 100% for human relations, rater, and general management training, respectively.

The estimated true mean effect sizes were substantial (more than one half of one standard deviation) in all cases. For the overall analysis with objective results as criteria, as well as for the rater training content distribution, the estimated true effect size variances were also large. As a result of large estimated true variances, these latter two distributions had negative lower bound credibility values. Human relations training, however, had a sizeable positive lower bound credibility estimate (.63).

Table 3
Results for Content of Managerial Training on Objective Learning Criteria

Training areas	No. of *ES*s	No. of studies	N	$\overline{ES}$	V_{obs}	V_e	% of variance accounted for	$\hat{M}_\delta$	$\hat{V}_\delta$	90% C.V.
Overall areas[a]	77	22	8,280	.33	.517	.052	13.3	.38	.339	−.37
Human relations	33	8	2,281	.33	.208	.090	35.8	.41	.161	−.10
General management	17	6	4,530	.18	.447	.019	25.1	.21	.057	−.10
Problem solving	11	3	605	.16	1.005	.094	14.4	.17	.573	−.80
Motivation/values	12	3	450	.74	.141	.153	100.0	.85	0	.85

Note. V_{obs} is the observed effect size variance. V_e is sampling error variance. $\overline{ES}$ is the observed weighted mean according to sample size.
[a] The content areas of self-awareness and rater training were not included in separate breakdowns because only two or less studies in each area were found that used objective learning criteria. However, they were included in the overall results for objective learning criteria if data were available.

Table 4
Results for Content of Managerial Training on Subjective Behavior Criteria

Training areas	No. of *ESs*	No. of studies	*N*	$\overline{ES}$	V_{obs}	V_e	% of variance accounted for	$\hat{M}_\delta$	$\hat{V}_\delta$	90% C.V.
Overall areas[a]	277	39	26,025	.44	.384	.055	13.8	.49	.344	−.26
Human relations	118	17	6,537	.39	.158	.094	60.5	.44	.061	.12
General management	88	11	11,707	.36	.476	.039	7.1	.40	.506	−.51
Self-awareness	52	7	6,944	.61	.837	.038	10.1	.65	.342	−.09

Note. V_{obs} is the observed effect size variance. V_e is sampling error variance. $\overline{ES}$ is the observed weighted mean according to sample size.
[a] The content areas of problem solving, motivation/values and rater training were not included in separate breakdowns because only two or less studies in each area were found that used subjective behavior criteria. However, they were included in the overall results for subjective behavior criteria if data were available.

Training Method

Subjective learning criteria. For the four training method distributions that had subjective learning criteria (see Table 6), artifactual sources of variance did not account for a large proportion of the observed effect size variance. The largest percentage of observed effect size variance accounted for was 39% for behavioral modeling. The estimated true mean effect sizes for the four training method distributions were relatively large in all cases. Also, although there was a fair degree of variation in the estimated true effects for these four distributions, all lower bound credibility value estimates were positive. Furthermore, behavioral modeling had a sizeable (.76) lower bound credibility value.

Objective learning criteria. As seen in Table 7, criterion unreliability and sampling error variance did not account for a large percentage of the observed effect size variances for the four training method distributions that had objective learning criteria. A majority of the observed effect size variance (56%) was accounted for by artifactual sources for the category of lecture with discussion and either role playing or practice.

The magnitude of the estimated true mean effect sizes for the four training method distributions varied considerably. Particularly notable was the size of the estimated true mean effect size of .93 for lecture with discussion and either role playing or practice. Moreover, the latter method category had a lower bound credibility value of .46. On the other hand, the lower bound credibility values for the remaining three training method categories were negative or approximately equal to zero.

Subjective behavior criteria. The managerial training method results for studies employing subjective measures of behavior as criteria are presented in Table 8. For three managerial training method categories—Leader Match, behavioral modeling, and lecture with discussion and either role playing or practice—a substantial percentage (70% or greater) of the observed effect size variance was accounted for by artifactual sources. In addition, the training method distributions for lecture and lecture with discussion had 47% and 51% of the observed effect size variance accounted for by artifacts, respectively. Sensitivity training and the multiple techniques distribution had small percentages of observed effect size variance accounted for by artifacts.

The estimated true mean effect sizes were positive and greater than or equal to .4 for five of the seven training method distributions. In particular, sensitivity training and behavioral modeling had estimated true mean effect sizes of .73 and .78, respectively. Sensitivity training, however, as well as the multiple techniques distribution had negative lower bound credibility values. The lower bound credibility values for lecture, lecture with discussion, and lecture with discussion and either role playing or practice were positive.

Objective results criteria. Two distributions of training method effects with objective results as criteria were analyzed and are reported in Table 9. The multiple techniques category had 100% of the observed effect size variance accounted for by artifactual sources of variance. The lecture method only had approximately 22% of the observed effect size variance accounted for by artifacts. In addition, the estimated true mean effect sizes were relatively large for both training method distributions. However, because of the large estimated true effect size variance for the lecture method, this method had a negative lower bound credibility value.

Table 5
Results for Content of Managerial Training on Objective Results Criteria

Training areas	No. of *ESs*	No. of studies	*N*	$\overline{ES}$	V_{obs}	V_e	% of variance accounted for	$\hat{M}_\delta$	$\hat{V}_\delta$	90% C.V.
Overall areas[a]	60	11	2,298	.57	.407	.110	22.0	.67	.323	−.06
Human relations	3	3	314	1.01	.349	.045	30.8	1.04	.102	.63
General management	10	4	606	.49	.110	.075	100.0	.53	0	.53
Rater training	46	3	1,326	.50	.478	.256	33.4	.64	.510	−.27

Note. V_{obs} is the observed effect size variance. V_e is sampling error variance. $\overline{ES}$ is the observed weighted mean according to sample size.
[a] The content area of Self-Awareness, Problem Solving and Motivation/Values were not included in separate breakdowns since only two or less studies in each area were found that used objective results criteria. However, they were included in the overall results for objective results criteria if data were available.

Table 6
Results for Training Methods on Subjective Learning Criteria

Method[a]	No. of *ES*s	No. of studies	N	$\overline{ES}$	V_{obs}	V_e	% of variance accounted for	$\hat{M}_\delta$	$\hat{V}_\delta$	90% C.V.
Sensitivity training	15	7	832	.80	.524	.091	18.6	.86	.399	.06
Behavioral modeling	7	3	657	.92	.305	.053	39.3	.99	.081	.76
Lecture/discussion plus role play or practice	10	3	1,058	.60	.218	.048	25.0	.66	.146	.17
Multitechnique (3 or more)	15	4	1,258	.70	.286	.060	24.4	.76	.185	.21

Note. V_{obs} is the observed effect size variance. V_e is sampling error variance. $\overline{ES}$ is the observed weighted mean according to sample size.
[a] The training methods of lecture, lecture/discussion and Leader Match were not included in separate breakdowns because only two or less studies for each method were found that used subjective learning criteria.

Discussion

Training Content

For subjective learning criteria, over all studies and for the specific content breakdowns (i.e., human relations, general management functions, and self-awareness), it was found that controllable statistical artifacts account for small to moderate percentages of the observed effect size variances. It is highly unlikely that additional, uncontrolled statistical artifacts would have substantially increased the percentages of observed variance accounted for. Relatively larger amounts of unaccounted-for variance in these distributions may well be explained by other substantive variables (e.g., training method) that we were not able to simultaneously control in this study. Furthermore, although care should be exercised in the selection and implementation of a human relations training program because the degree of its effectiveness is likely to vary with other substantive variables, the trainer (and organization) can be fairly confident that this type of training will be effective in improving subjective learning.

The results for objective learning criteria, over all training studies, indicated that controlled-for statistical artifacts accounted for only a small percentage of the observed effect size variance. In addition, the breakdowns for three of the four training content areas (i.e., human relations, general management, and problem solving) did not provide plausible explanations for this large unaccounted-for variation in effect sizes.

The highly negative results for problem-solving and decision-making training were based on three studies with a total of 11 effects. As early as 1970, Campbell et al. stated disappointment with the results and number of studies evaluating problem-solving and decision-making training. Because of the lack of studies for this analysis, as well as the incompleteness of the reported results, the effectiveness of problem-solving and decision-making training programs is still difficult to estimate.

In addition, the small true mean effect size for general management involving objective learning criteria was probably due to a study in which a large sample was used, and few significant changes were found on the knowledge tests (Goodacre, 1957). In another study evaluating learning criteria in general management training (Mahoney, Jerdee, & Korman, 1960), the trained group actually had lower scores than did the control group on the knowledge tests. To explain these results, the authors claimed that the criterion measure used was not relevant to the training program and that instructors were not adequately prepared. Because there have been few well-designed studies within this area, no definite conclusions can be reached as to the effectiveness of general management functions training.

It was found, however, that training programs that focused on increasing motivation or improving values as measured by objective learning criteria were quite effective. That is, managers who received training were moved to the 79th percentile of the untrained group. For this analysis, the situational specificity hy-

Table 7
Results for Training Methods on Objective Learning Criteria

Method[a]	No. of *ES*s	No. of studies	N	$\overline{ES}$	V_{obs}	V_e	% of variance accounted for	$\hat{M}_\delta$	$\hat{V}_\delta$	90% C.V.
Lecture	20	5	1,708	.28	.134	.078	43.6	.37	.101	−.03
Lecture/discussion	24	4	4,782	.20	.202	.027	34.1	.23	.051	−.06
Lecture/discussion plus role play or practice	8	3	267	.82	.345	.172	55.8	.93	.136	.46
Multitechnique (3 or more)	13	6	607	.73	.609	.117	23.3	.81	.385	.01

Note. V_{obs} is the observed effect size variance. V_e is sampling error variance. $\overline{ES}$ is the observed weighted mean according to sample size.
[a] The training methods of Leader Match, sensitivity training and behavioral modeling were not included in separate breakdowns because only two or less studies for each method were found that used objective learning criteria.

Table 8
Results for Training Methods on Subjective Behavior Criteria

Method	No. of *ES*s	No. of studies	N	$\overline{ES}$	V_{obs}	V_e	% of variance accounted for	$\hat{M}_\delta$	$\hat{V}_\delta$	90% C.V.
Lecture	12	3	1,055	.41	.076	.059	47.4	.46	.065	.13
Lecture/discussion	11	4	5,102	.10	.115	.012	50.9	.11	.012	.03
Leader Match	69	5	3,081	.36	.137	.117	100.0	.40	0	.40
Sensitivity training	49	8	7,153	.67	.712	.035	7.8	.73	.453	−.13
Behavioral modeling	17	5	446	.70	.275	.201	100.0	.78	0	.78
Lecture/discussion plus role play or practice	21	4	1,117	.30	.169	.096	70.0	.34	.042	.07
Multitechnique (3 or more)	76	11	5,169	.45	.419	.078	11.7	.51	.597	−.48

Note. V_{obs} is the observed effect size variance. V_e is sampling error variance. $\overline{ES}$ is the observed weighted mean according to sample size.

pothesis can be rejected because 100% of the observed effect size variance was accounted for by artifactual effects. Therefore, managerial training for change in motivation and values does appear to lead to increased motivation and appropriate value changes as measured by objective learning criteria. However, these results should be interpreted with some caution, considering that only three studies with a total of 12 effects were found in this area. Although the effectiveness of managerial training as measured by learning criteria is important, the effects of training on job performance and behavior is of considerable importance to the individual manager and to the organization.

As with learning criteria, after correcting for the artifactual effects of sampling error and criterion reliability in the distributions with subjective behavior criteria, a substantial amount of observed effect size variance remained unaccounted for in the overall analysis, as well as in the specific content breakdowns. A primary factor influencing this overall result was the 11 studies for general management training that had a lower bound credibility estimate of −.51. If other artifacts had been removed, such as range restriction, the resulting true variance would have been even smaller. It is highly unlikely, however, that accounting for other statistical artifactual sources would have altered this result. As noted above, a more likely explanation is that other substantive variables moderate the effectiveness of different types (i.e., content) of managerial training.

In contrast to the results based on subjective and objective learning criteria, the magnitudes of the true mean effect sizes based on subjective behavior criteria were similar for the training content areas. That is, the range of true mean effect size was .40 to .65. In particular, self-awareness training, which typically uses some form of laboratory education/sensitivity training, was also shown to be fairly effective, on the average, in changing managerial behavior on the job. Although self-awareness training yielded the highest true mean effect size, the criterion measures typically used for evaluating self-awareness training were of questionable utility. In many instances, observers were asked to report changes in on-the-job behavior, or individual self-reports of behavior change were used as an index of training effectiveness. These criterion measures may not necessarily reflect changes in actual job performance. Therefore, it is suggested that in the future, evaluations of self-awareness training use actual job performance criteria.

The results of the overall analysis for objective results criteria indicated that a majority of the observed variation in effect sizes was not explained by statistical artifacts. The results for the content area of human relations training, however, revealed this type of training to be on the average very effective and likely to produce some improvement in managerial performance regardless of the situation. It should be noted that only three studies were involved in this analysis and thus the power of this analysis is suspect.

Similarly, some caution is warranted in the interpretation of the rater training results because only three studies were examined. More research using management personnel as subjects is needed before more definitive conclusions can be drawn regarding the effectiveness of rater training with respect to job performance for these individuals.

In contrast to the results based on learning criteria and subjective behavior, general management training was shown to be

Table 9
Results for Training Methods on Objective Results Criteria

Method[a]	No. of *ES*s	No. of studies	N	$\overline{ES}$	V_{obs}	V_e	% of variance accounted for	$\hat{M}_\delta$	$\hat{V}_\delta$	90% C.V.
Lecture	15	3	520	.64	.394	.125	21.5	.82	.456	−.04
Multitechnique (3 or more)	13	5	634	.49	.099	.096	100.0	.52	0	.52

Note. V_{obs} is the observed effect size variance. V_e is sampling error variance. $\overline{ES}$ is the observed weighted mean according to sample size.
[a] The training methods of Lecture/Discussion, Leader Match, Sensitivity Training, Behavioral Modeling and Lecture/Discussion plus Role Play or Practice were not included in separate breakdowns since only two or less studies for each method were found that used objective results criteria.

very effective, on the average, in improving performance as measured by objective results. Based on the finding that this training was shown to generalize across settings for objective results criteria, the potential for general management training's having an impact on organization's bottom line in a new setting may be high. More research in this area would be enlightening, especially from the viewpoint of evaluating the influence of general management training on organizational economic gains. Although training content such as general management is an important factor in choosing a training program for a particular application and in providing a means of organizing evaluations of training programs, another important factor is the method of training.

Training Method

The results for the four training method categories with respect to subjective learning criteria indicated that these assisted to a moderate degree in explaining the unaccounted-for observed effect size variance for subjective learning criteria. It is important to note that the methods of behavioral modeling, sensitivity training, lecture with discussion and either role playing or practice, and multiple techniques are highly likely to lead to positive training results in a new situation at least at a minimal level. In particular, the results suggested that behavioral modeling is a sound method for improving learning across situations as measured by subjective learning criteria. Overall, these are encouraging results for the aforementioned training methods in regard to subjective learning criteria.

For studies using objective learning criteria, the results indicated that the training methods of lecture, lecture with discussion, and lecture with discussion and either role playing or practice did assist in accounting for some of the unexplained variance for the overall analysis with objective learning criteria. An important finding was that a majority (56%) of the observed effect size variance for the category of lecture with discussion and either role playing or practice was accounted for by criterion unreliability and sampling error. Furthermore, this latter category had a sizeable lower bound credibility value, indicating that this method of training is very likely to generalize across situations using objective learning criteria.

The results for studies using subjective behavior criteria indicated that unaccounted-for variance in the overall analysis can partially be explained by training method. That is, the lecture, lecture with discussion, lecture with discussion and either role playing or practice, Leader Match, and behavioral modeling methods were helpful in explaining the large amount of unaccounted-for variance in the overall analysis for subjective behavior criteria. Noteworthy were the positive lower bound credibility values obtained for the three distributions of studies using the lecture method. These results indicate that training that employs the lecture method is likely to generalize across situations to some degree. These positive results for the lecture method are encouraging in light of authors (Bass & Vaughn, 1966; Korman, 1977; McGehee & Thayer, 1961) who have questioned the usefulness of this method and training directors who have expressed low opinions of the lecture method (cf. Caroll, Paine, & Ivancevich, 1972).

The results suggest that the effectiveness of the Leader Match training method with respect to subjective behavior criteria generalizes across situations. The results for the Leader Match training method are consistent with the findings of Rice (1978, 1979) and Fiedler and Mahar (1979a, 1979b). On the basis of these results, as well as the cost-effectiveness of Leader Match training compared with that of other leadership training programs, this method of leadership training is encouraged.

The effectiveness of managerial behavioral modeling training with respect to subjective behavior criteria was also shown to generalize across settings. This finding is consistent with the impressive empirical support for social learning theory obtained from well-controlled studies in experimental situations (cf. Bandura, 1977), as well as previous findings in organizational settings (cf. Burnaska, 1976; Byham, Adams, & Kiggins, 1976; Latham & Saari, 1979; Smith, 1976). The magnitude of the estimated true mean effect for behavioral modeling provides an indication of how useful this method of managerial training is likely to be in improving managerial behaviors.

Another interesting finding for studies using subjective behavior criteria was that managerial sensitivity training had a relatively high true mean effect size. This result is by definition consistent with the breakdown of training content for self-awareness training. That is, sensitivity training was always categorized as self-awareness training for training content and likewise for training method. Consequently, when one of these variables (i.e., content or method) was controlled, the other, of necessity, was also controlled. These results, however, do not necessarily point to situational specificity of sensitivity training. As with other breakdowns, other substantive variables (e.g., qualifications of the trainer) may help explain some of the large unaccounted-for variance in observed effect sizes. It is unlikely that substantive variables would assist in explaining most of this large unaccounted-for variation. These results are not encouraging for the generalizability of sensitivity training with respect to subject behavior criteria; future research will be necessary to make such determinations. Furthermore, these findings are in contrast to Smith's (1975) conclusions regarding the effectiveness of sensitivity training. The typical criterion measures used for evaluating sensitivity training were also of questionable utility. In many instances, either observers were asked to report changes in on-the-job behavior, or individual self-reports of behavior changes were used as an index of training effectiveness. These criterion measures may not necessarily reflect changes in actual job performance. For a discussion of some of the potential problems with sensitivity training, the reader is referred to Wexley and Latham (1981).

The results for one of the two distributions analyzed for objective results criteria, the multiple training techniques distribution, tended to generalize across settings. The theoretical and practical soundness of this latter result is questionable because a variety of different techniques were included in each study. In addition, the statistical power of this latter finding is suspect considering that the analogous finding for subjective behavior criteria had twice as many studies and that approximately six times the number of effects and different results were obtained. Overall, too few studies, for all methods of managerial training, used objective results criteria necessary to conduct meaningful

analyses and derive sound conclusions with respect to these types of dependent variables.

Limitations

Some limitations of the present meta-analysis should be noted. The number of studies was relatively small. Although an attempt was made to locate all relevant studies, it is likely that some pertinent studies were not included. The relatively small number of studies, however, is more a reflection of the lack of empirical research on managerial training than of the thoroughness of the present literature search. Another limitation is that only two artifactual sources, criteron reliability and sampling error, were taken into account. Other potentially important artifactual sources of variance (e.g., range restriction, typographical errors) could not be controlled or corrected for. By not correcting for other sources of artifactual variance, the variance due to situational specificity was slightly inflated. The degree of inflation was not evaluated. A third limitation is that criterion reliability data was not available in many instances and had to be estimated. The degree to which the estimated reliabilities reflected the actual values was undeterminable. This study also suffered from the inability to conduct some potentially useful follow-up analyses because many studies did not report necessary information.

Conclusions

One conclusion is that researchers need to improve their reports evaluating organizational interventions such as providing information on the degree of range restriction, criterion and predictor reliabilities, sample characteristics, and a thorough description of their methodology. Such information will equip investigators with the necessary data to perform cumulative analyses of the effectiveness of managerial training, as well as that of other organizational interventions.

Another conclusion, based on the results for the training content distributions, is that trainers and organizational decision makers should not rely heavily on training program content descriptions and labels when choosing and judging the probable utility of a managerial training program. More important, the results from the training method analyses point toward the choice of a particular method that might be most effective in improving results related to a certain type of criterion measure. To aid in such decision making, further research aimed at determining the effectiveness of different training methods with respect to different types of dependent variables appears worthwhile. In particular, the breakdowns in Tables 6 through 9 with relatively few studies (i.e., five or less), as well as the breakdowns that could not be reported because of insufficient number of studies (e.g., lecture method on subjective learning criteria), suggest areas where there is a pressing need for sound empirical research.

In addition, the level of experience of the trainer may be a significant factor influencing the effectiveness of training programs. Future research aimed at assessing the influence of the trainer's experience and qualifications on the effectiveness of training would be enlightening.

Moreover, the results of this meta-analysis indicate that different managerial training methods do not necessarily lead to increased knowledge and improved job performance. These findings are consistent with previous reviews of the managerial training literature (Campbell, 1971; Campbell et al., 1970; Goldstein, 1980). The results from this meta-analysis go beyond these reviews by quantitatively evaluating the degree to which the effectiveness of managerial training generalizes across settings for various training content areas, training methods, and outcome measures. Overall, different methods of managerial training are on the average moderately effective in improving learning and job performance. Furthermore, in most instances, positive lower bound credibility values were obtained, and these effects do indicate gains in knowledge and improved performance. Even small effects of less than one half of one standard deviation have been shown through utility analysis to lead to a substantial economic impact on the organization (Hunter & Schmidt, 1983). It is hoped that this study will aid researchers and practitioners in judging the probable utility of different managerial training programs, as well as alert them to the need to empirically evaluate and thoroughly report the effectiveness of organizational interventions.

In conclusion, this meta-analysis can be viewed as an initial step in clarifying what we have learned about managerial training and highlighting areas where future research is needed. The completion of well-designed, thoroughly reported empirical studies in the areas noted above will provide the necessary data for more refined meta-analysis of managerial training and training in general.

References

Alves, E., & Hardy, W. R. (1963). Evaluating supervisory training in Los Angeles County. *Training Directors, 17,* 36–40.

Argyris, C. (1965). Explorations in interpersonal competence—II. *Journal of Applied Behavioral Science, 1,* 255–269.

Arnoff, J., & Litwin, G. H. (1971). Achievement motivation training and executive advancement. *Journal of Applied Behavioral Science, 7,* 215–229.

Bandura, S. (1977). *Social learning theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bass, B. M., & Vaughn, J. A. (1966). *Training in industry: The management of learning.* Belmont, CA: Wadsworth.

Baum, B. H., Sorensen, P. F., Jr., & Place, W. S. (1970). The effect of managerial training on organizational control: An experimental study. *Organizational Behavior and Human Performance, 5,* 170–182.

Boyd, J. B., & Ellis, J. D. (1962). *Findings of research into senior management seminars.* Toronto: Hydro-Electric Power Commission of Ontario.

Buchanan, P. C., & Brunstetter, P. H. (1959). A research approach to management improvement. *Journal of the American Society of Training Directors, 13,* 18–28.

Bunker, D. R. (1965). Individual applications of laboratory training. *Journal of Applied Behavioral Science. 1,* 131–147.

Burnaska, R. F. (1976). The effects of behavior modeling and training upon managers' behaviors and employees' perceptions. *Personnel Psychology, 29,* 329–335.

Byham, W. C., Adams, D., & Kiggins, A. (1976). Transfer of modeling training to the job. *Personnal Psychology, 29,* 345–349.

Campbell, J. P. (1971). Personnel training and development. *Annual Review of Psychology, 22,* 565–602.

Campbell, J. P., Dunnette, M. D., Lawler, E. E., & Weick, K. E., Jr.

(1970). *Managerial behavior, performance, and effectiveness.* New York: McGraw-Hill.

Canter, R. R., Jr. (1951). A human relations training program. *Journal of Applied Psychology, 35,* 38–45.

Caroll, S. J., Paine, E. T., & Ivancevich, J. J. (1972). The relative effectiveness of training methods—expert opinion and research. *Personnel Psychology, 25,* 495–510.

Carron, T. J. (1964). Human relations training and attitude change: A vector analysis. *Personnel Psychology, 17,* 403–424.

Couch, P. D., & Strother, G. B. (1971). A critical incident evaluation of supervisory training. *Training and Development Journal, 25,* 6–11.

Csoka, L. S., & Bons, P. M. (1978). Manipulating the situation to fit the leader's style: Two validation studies of Leader Match. *Journal of Applied Psychology, 63,* 295–300.

Decker, P. (1982). The enhancement of behavior modeling training of supervisory skills by the inclusion of retention processes. *Personnel Psychology, 35,* 323–331.

DiVesta, F. J. (1954). Instruction-centered and student-centered approaches in teaching a human relations course. *Journal of Applied Psychology, 38,* 329–335.

Durand, D. E. (1975). Effects of achievement motivation and skill training on the entrepreneurial behavior of black businessmen. *Organizational Behavior and Human Performance, 14,* 76–90.

Fiedler, F. E. (1964). A contingency model of leadership effectiveness. In Berkowitz, L. (Ed.), *Advances in experimental psychology* (pp. 149–190). New York: Academic Press.

Fiedler, F. E. (1967). *A theory of leadership effectiveness.* New York: McGraw-Hill.

Fiedler, F. E., Chemers, M. M., & Mahar, L. (1976). *Improving leadership effectiveness: The Leader Match concept.* New York: Wiley.

Fiedler, F. E., & Mahar, L. (1979a). The effectiveness of contingency model training: A review of the validation of Leader Match. *Personnel Psychology, 32,*45–62.

Fiedler, F. E., & Mahar, L. (1979b). A field experiment validating contingency model leadership training. *Journal of Applied Psychology, 64,* 247–254.

Fiedler, F. E., Mahar, L. & Schmidt, D. (1976). *Four validation studies of contingency model training* (Tech. Rep. No. 75–70). Seattle, WA: Organizational Research, University of Washington.

Fleishman, E. A. (1953). Leadership climate, human relations training, and supervisory behavior. *Personnel Psychology, 6,* 205–222.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

Goldstein, A. P., & Sorcher, M. (1974). *Changing supervisor behavior.* New York: Pergamon Press.

Goldstein, I. L. (1980). Training in work organizations. *Annual Review of Psychology, 31,* 229–272.

Goodacre, D. M. (1957). The experimental evaluation of management training principles and practice. *Personnel, 33,* 534–538.

Goodacre, D. M. (1963). Stimulating improved man management. *Personnel Psychology, 16,* 133–143.

Gruenfeld, L. W. (1966). Management development effect on changes in values. *Training and Development Journal, 20,* 18–26.

Hand, H. H., Richard, M. D., & Slocum, J. W., Jr. (1973). Organizational climate and the effectiveness of a human relations training program. *Academy of Management Journal, 16,* 185–195.

Hand, H. H., & Slocum, J. W., Jr. (1970). Human relations training for middle management: A field experiment. *Academy of Management Journal, 13,* 403–410.

Hand, H. H., & Slocum, J. W., Jr. (1972). A longitudinal study of the effects of a human relations training program on managerial effectiveness. *Journal of Applied Psychology, 56,* 412–417.

Harris, E. F., & Fleishman, E. A. (1955). Human relations training and the stability of leadership patterns. *Journal of Applied Psychology, 39,* 20–25.

Harrison, R. (1962). Impact of the laboratory on perceptions of others by the experimental group. In C. Argyris (Ed.), *Interpersonal competence and organizational effectiveness* (pp. 261–271). Homewood, IL: Dorsey-Irwin.

Harrison, R., & Lubin, B. (1965). Personal style, group composition, and learning. *Journal of Applied Behavioral Science, 1,* 286–301.

Hinrichs, J. R. (1976). Personnel training. In M. D. Dunnette, (Ed.), *Handbook of industrial and organizational psychology* (pp. 829–860). Chicago: Rand McNally.

House, R. J., & Tosi, H. (1963). An experimental evaluation of a management training program. *Academy of Management Journal, 6,* 303–315.

Hunter, J. E., & Schmidt, F. L. (1983). Quantifying the effects of psychological interventions on employee job performance and work-force productivity. *American Psychologist, 38,* 473–478.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-Analysis.* Beverly Hills, CA: Sage.

Ivancevich, J. M. (1979). Longitudinal study of the effects of rater training on psychometric error in ratings. *Journal of Applied Psychology, 64,* 502–508.

Ivancevich, J. M., & Smith, S. V. (1981). Goal setting interview skills training: Simulated and on-the-job analyses. *Journal of Applied Psychology, 66,* 697–705.

Jennings, E. E. (1954). The dynamics of forced leadership training. *Journal of Personnel Administration and Industrial Relations, 1,* 110–118.

Katz, S. I., & Schwebel, A. I. (1976). The transfer of laboratory training. *Small Group Behavior, 7,* 271–285.

Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), *Training and Development Handbook* (2nd ed., Chapter 18). New York: McGraw-Hill.

Kondrasuk, J. N. (1972). *Evaluation of management-by-objectives training.* Unpublished doctoral dissertation, University of Minnesota.

Korman, A. K. (1977). *Organizational behavior.* Englewood Cliffs, NJ: Prentice-Hall.

Latham, G. P., & Kinne, S. B. III. (1974). Improving job performance through training in goal setting. *Journal of Applied Psychology, 59,* 187–191.

Latham, G. P., & Saari, L. M. (1979). The application of social learning theory to training supervisors through behavior modeling. *Journal of Applied Psychology, 64,* 239–246.

Lawshe, C. H., Bolda, R. A., & Brune, R. L. (1959). Studies in management training evaluation: II. The effects of exposures to role playing. *Journal of Applied Psychology, 43,* 287–292.

Leister, A., Borden, D., & Fiedler, F. E. (1977). Validation of contingency model leadership training: Leader Match. *Academy of Management Journal, 20,* 464–470.

Levine, J., & Butler, J. (1952). Lecture versus group decision in changing behavior. *Journal of Applied Psychology, 36,* 29–33.

Lundberg, C., Dunbar, R., & Bayless, T. L. (1973). Contemporary management training in large corporations. *Training and Development Journal, 27,* 34–38.

Mahar, C. A. (1981). Training of managers in program planning and evaluation: Comparison of two approaches. *Journal of Organizational Behavior Management, 3,* 45–56.

Mahoney, T. A., Jerdee, T. H., & Korman, A. (1960). An experimental evaluation of management development. *Personnel Psychology, 13,* 147–163.

Maier, N. R. F. (1953). An experimental test of the effect of training on discussion leadership. *Human Relations, 6,* 161–173.

Mayo, G. D., & DuBois, P. H. (1963). Measurement of gain in leadership training. *Educational and Psychological Measurement, 23,* 23–31.

McGehee, W., & Gardner, J. E. (1955). Supervisory training and attitude change. *Personnel Psychology, 8,* 449–460.

McGehee, W., & Thayer, P. W. (1961). *Training in business and industry.* New York: Wiley.

Miles, M. B. (1960). Human relations training: Processes and outcomes. *Journal of Counseling Psychology, 7,* 301–306.

Miles, M. B. (1965). Changes during and following laboratory training: A clinical-experimental study. *Journal of Applied Behavioral Science, 1,* 215–242.

Miner, J. B. (1960). The effect of a course in psychology on the attitudes of research and development supervisors. *Journal of Applied Psychology, 44,* 224–232.

Miner, J. B. (1965). *Studies in management education.* New York: Springer.

Moffie, D. J., Calhoon, R., & O'Brien, J. K. (1964). Evaluation of a management development program. *Personnel Psychology, 17,* 431–440.

Moon, G. G., & Hariton, T. (1958). Evaluating an appraisal and feedback training program. *Personnel, 35,* 36–41.

Mosel, J. N., & Tsacnaris, H. J. (1954). Evaluating the supervisor training program. *Journal of Personnel Administration and Industrial Relations, 1,* 99–104.

Moses, J. L., & Ritchie, R. J. (1976). Supervisory relationships training: A behavioral evaluation of a behavior modeling program. *Personnel Psychology, 29,* 337–343.

Reith, J. (1976). Group methods: Conferences, meetings, workshops, seminars. In Craig, R. L. (Ed.), *Training and development handbook* (2nd ed., Chapter 34). New York: McGraw-Hill.

Rice, R. W. (1978). Construct validity of the Least Preferred Co-Worker (LPC) score. *Psychological Bulletin, 82,* 597–622.

Rice, R. W. (1979). Reliability and validity of the LPC scale: A reply. *Academy of Management, 4,* 291–294.

Roy, S. K., & Dolke, A. M. (1971). Evaluation of a supervisory training program. *Training and Development Journal, 25,* 35–39.

Rubin, I. (1967). Increased self-acceptance: A means of reducing social prejudice. *Journal of Personality and Social Psychology, 5,* 233–239.

Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1982). Assessing the economic impact of personnel programs on workforce productivity. *Personnel Psychology, 35,* 333–347.

Schutz, W. C., & Allen, V. L. (1966). The effects of a T-group laboratory on interpersonal behavior. *Journal of Applied Behavioral Science, 3,* 265–286.

Schwarz, F. D., Stilwell, W. P., & Scanlan, B. K. (1968). Effects of management development on manager behavior and subordinate perception: Part I. *Training and Development Journal, 22,* 38–50.

Smith, E. E., & Kight, S. S. (1959). Effects of feedback on insight and problem solving efficiency in training groups. *Journal of Applied Psychology, 43,* 209–211.

Smith, P. B. (1963). Attitude changes associated with training in human relations. *British Journal of Social and Clinical Psychology, 2,* 104–112.

Smith, P. B. (1975). Controlled studies of the outcomes of sensitivity training. *Psychological Bulletin, 82,* 597–622.

Smith, P. E. (1976). Management modeling training to improve morale and customer satisfaction. *Personnel Psychology, 29,* 351–359.

Steele, F. I. (1968). Personality and the "laboratory style." *Journal of Applied Behavioral Science, 4,* 25–45.

Stroud, P. V. (1959). Evaluating a human relations training program. *Personnel, 36,* 52–60.

Trice, H. M., & Belasco, J. A. (1968). Supervisory training about alcoholics and other problem employees. *Quarterly Journal of Studies on Alcohol, 29,* 382–398.

Underwood, W. J. (1965). Evaluation of laboratory-method training. *Training Directors Journal, 19,*34–40.

Valiquet, M. I. (1968). Individual change in a management development program. *Journal of Applied Behavioral Science, 4,* 313–326.

Viteles, M. S. (1959). Human relations and the humanities in the education of business leaders: Evaluation of a program of humanistic studies for executives. *Personnel Psychology, 12,* 1–28.

Warmke, D. L., & Billings, R. S. (1979). Comparison of training methods for improving the psychometric quality of experimental and administrative performance ratings. *Journal of Applied Psychology, 64,* 124–131.

Wexley, K. N. (1984). Personnel training. *Annual Review of Psychology, 35,* 519–551.

Wexley, G. P., & Latham, K. N. (1981). *Developing and training human resources in organizations.* New York: Scott, Foresman and Company.

Wexley, K. N., & Nemeroff, W. F. (1975). Effectiveness of positive reinforcement and goal setting as methods of management development. *Journal of Applied Psychology, 60,* 446–450.

Wikstrom, W. S. (1973). *Supervisory training* (Rep. No. 612). New York Conference Board.

Wilburn, J. M. III. (1979). *An evaluation study: The relationship of cognitive and affective perceptions to participation in a management development program.* Unpublished doctoral dissertation, North Carolina State University.

Wilson, J. E., Mullen, D. P., & Morton, R. B. (1968). Sensitivity training for individual growth-team training for organization development. *Training and Development Journal, 22,* 47–53.

Wolfe, J., & Moe, B. L. (1973). An experimental evaluation of a hospital supervisory training program. *Hospital Administration, 18,* 65–77.

Received November 26, 1984
Revision received June 10, 1985 ■